Minimum spanning trees on weighted scale-free networks

P. J. Macdonald, E. Almaas, and A.-L. Barabási*
Center for Network Research and Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Dated: February 2, 2008)

A complete understanding of real networks requires us to understand the consequences of the uneven interaction strengths between a system's components. Here we use the minimum spanning tree (MST) to explore the effect of weight assignment and network topology on the organization of complex networks. We find that if the weight distribution is correlated with the network topology, the MSTs are either scale-free or exponential. In contrast, when the correlations between weights and topology are absent, the MST degree distribution is a power-law and independent of the weight distribution. These results offer a systematic way to explore the impact of weak links on the structure and integrity of complex networks.

PACS numbers: 89.75.Hc,89.75.Fb,89.75.Da,05.10.-a

The study of many complex systems have benefited from representing them as networks [1], examples including metabolic networks [2], describing the reactions in a cell's metabolism; the protein interaction network [3], capturing the binding interactions between a cell's proteins; and the World Wide Web and email networks [4, 5] linking web-pages or people together via URLs or emails. For these systems there is extensive empirical evidence indicating that the degree (or connectivity) distribution of the nodes follows a power-law, strongly influencing everything from network robustness [6] to disease spreading [7]. However, to fully characterize these systems, we need to acknowledge the fact that the links can differ in their strength and importance [8, 9, 10, 11, 12]. Indeed, in a social network the strength of the relationship between two long-time friends differs from that between two casual business associates [13]; in ecological systems the strength of a particular pair-interaction between species is crucial for population dynamics [14], ecosystem stability [15] and development in stressed environments [16]. Thus in most networks the links are not binary (present or absent), but have a strength that quantifies the importance of the particular node-to-node interaction.

The weakest links can carry particular significance in some weighted networks[13]. For instance, the speed of data transmission between two computers is limited by the link with the smallest bandwidth ("bottleneck"), or the activity of a metabolic pathway is determined by the rate of the slowest reaction. Furthermore, weak links can affect the overall network integrity. For example, ecological communities may experience dramatic effects upon the removal of weak interactors [15]. To systematically uncover the location and the role of weak links in a complex network, we use the minimum spanning tree (MST), which for an N node network represents the loopless subgraph of $(N-1)$ links that reaches all nodes while *minimizing* the sum of the link weights [17, 18, 19, 20]. By avoiding the strong links and preferentially following the weakest ones, the MST selects the lowest weight backbone of a network.

We start by examining the correlations between weights and network structure for several real systems, allowing us to construct a model system whose weight distribution mimics the statistical features of real networks. We then show that the large-scale structure of the MSTs depends on the way the weights are placed in the network: For systems whose weight distribution is correlated with the network topology, the MSTs are either scale-free or exponential. In contrast, when the correlations between weights and topology are removed, the MST degree distribution is a power-law with a degree exponent close to the degree exponent of the original network, independent of the weight distribution.

Topology Correlated Weights.—To uncover the functional relationship between network topology and link weights, in Fig. 1 we display the dependence of the weights on the node degrees for the *E. coli* metabolic network, where the link weights represent the optimal metabolic fluxes [21]; the US Airport Network (USAN) where the weights reflect the total number of passengers travelling between two airports between 1992 and 2002; and the link betweenness-centrality (BC), representing the number of shortest paths along a link for the Barabási-Albert (BA) scale-free model [22]. For each of these systems the weight distributions follow a power-law [9] (not shown) and, as Fig 1 shows, the average link weight scales with the degrees of the nodes on the two ends of a link as $\langle w_{ij} \rangle \sim (k_i k_j)^{\theta}$, similar to the scaling found for the World Airport Network [11].

These empirical observations allow us to assign weights to the links of a network for which we have only the network topology. To systematically study the role of the weight distribution on the structure of the MST we use several weight assignments. (i) First, we choose $w_{ij} = k_i k_j$ (see inset Fig. 1). Note that the MST generated by this weight assignment is identical to the MST obtained for weights $w'_{ij} = (w_{ij})^{\theta}$ with any $\theta > 0$, as it is the rank of the weights and not their absolute value that

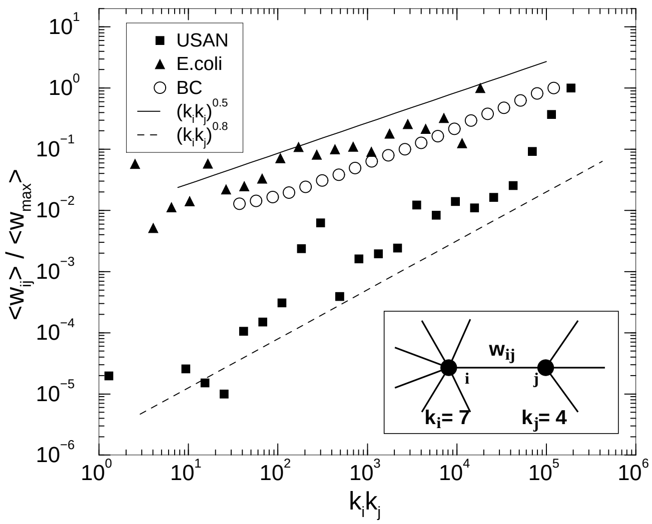

FIG. 1: The average weight of a link between nodes i and j shown as function of the link end-point degree product k_ik_j. The symbols represent (i) the USAN with the number of passengers as link weights (filled squares); (ii) the *E. coli* metabolic network with optimized flux as link weight [21] (filled triangles); (iii) Barabási-Albert scale-free model with betweenness-centrality (BC) as link weights (open circles). The solid line ($w \sim (k_ik_j)^{0.5}$) and the dashed line ($w \sim (k_ik_j)^{0.8}$) serve as guides to the eye. **Inset:** The weights are determined by the end-point degrees k_i and k_j: (i) $w_{ij} = k_ik_j$, (ii) $w_{ij} = \max(k_i, k_j)$, (iii) $w_{ij} = \min(k_i, k_j)$ or the inverse thereof (iv)-(vi).

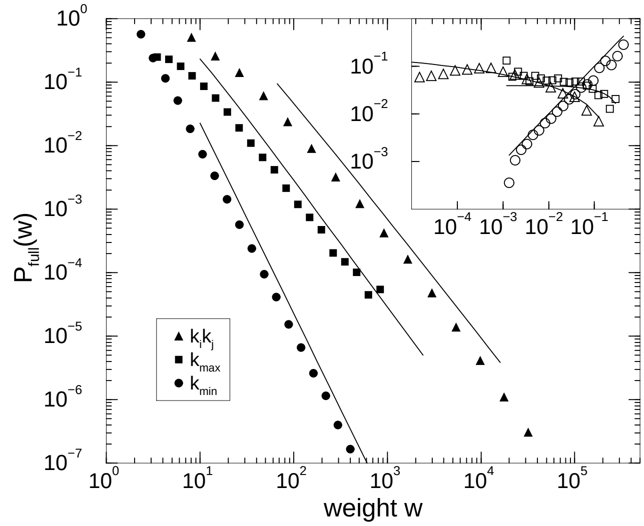

FIG. 2: Distribution of link weights on $N = 10^5$ node scale-free networks. Link-weight choice (i) (triangles), (ii) (squares) and (iii) (circles) are all heavy tailed. The analytical predictions (Eqs. (1) - (3)) are indicated as solid lines. Note that the solid curves have been shifted vertically without changing the character of the scaling law. **Inset:** The inverse weight distributions (iv)-(vi) (triangles, squares and circles respectively) and the analytical predictions shown as continuous lines.

determines the MST [23]. We have also studied the two extreme cases of topology-correlated weights, distributed according to (ii) $w_{ij} \sim k_{max}$ and (iii) $w_{ij} \sim k_{min}$, where $k_{min} = \min(k_i, k_j)$ and $k_{max} = \max(k_i, k_j)$ and with $k_{min}^2 \leq k_ik_j \leq k_{max}^2$. Finally, we investigated the structure of the maximal spanning trees [24] for the above weight choices by determining the MST after transforming cases (i)-(iii) as $w'_{ij} = 1/w_{ij}$, resulting in the link-weight choices (iv) $w_{ij} \sim 1/k_ik_j$, (v) $w_{ij} \sim 1/k_{max}$ and (vi) $w_{ij} \sim 1/k_{min}$.

Weight Distributions.—To characterize the obtained weighted networks we first study their weight distribution. For this, we grow scale-free networks according to the BA model [22], the resulting networks having a degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$. We then assign a weight to each link according to (i)-(vi). For networks whose degrees at the two ends of a link are uncorrelated we can determine the weight distribution analytically using order statistics [25], finding

$$P_{k_ik_j}(w) = (\gamma-2)^2 \; m^{2(\gamma-2)} \; w^{-\gamma+1} \; \ln(w/m^2), \quad (1)$$

$$P_{k_{max}}(w) = 2(\gamma-2) m^{\gamma-2} w^{-\gamma+1} \left[1 - \left(\frac{w}{m}\right)^{-\gamma+2}\right], \quad (2)$$

$$P_{k_{min}}(w) = 2(\gamma-2) m^{2(\gamma-2)} w^{-2\gamma+3}. \quad (3)$$

Corresponding expressions for the inverse degree correlations are obtained after the variable change $w' = 1/w$. For $w_{ij} \sim (k_ik_j)^\theta$ the relationship between the exponent of the weight distribution ($P(w) \sim w^{-\sigma}, w \gg 1$), the exponent of the degree distribution γ and θ is

$$\sigma = 1 + \frac{\gamma - 2}{\theta}, \quad (4)$$

valid for cases (i) and (ii). In Fig. 2 we compare the numerically determined weight distributions with the scaling predicted by our analytical expressions, finding that the numerical curves display a w-dependency close to that of Eqs. (1)-(3), unaffected by the degree-degree correlations in the model [26, 27]. Note that power-law weight distributions like those in Fig. 2 have been observed for a wide range of network based dynamical processes [28].

Minimum Spanning Trees.—The MSTs were generated using Prim's greedy algorithm [29]: starting from a randomly selected node, at each time step we add the link (and hence a node) with the smallest weight among the links connected to the already accepted nodes. Whenever m links with the same (smallest) weight are encountered, we break the degeneracy by randomly selecting one among them with probability $1/m$.

The numerical results indicate that the degree distribution of the resulting MSTs fall into two distinct classes [30, 31]. Weight choices (i) and (ii) give rise to exponential MST degree distributions (Fig. 3a), while choices (iii)-(vi) result in power-law distributed MST degrees (Fig. 3b). We can understand the exponential nature

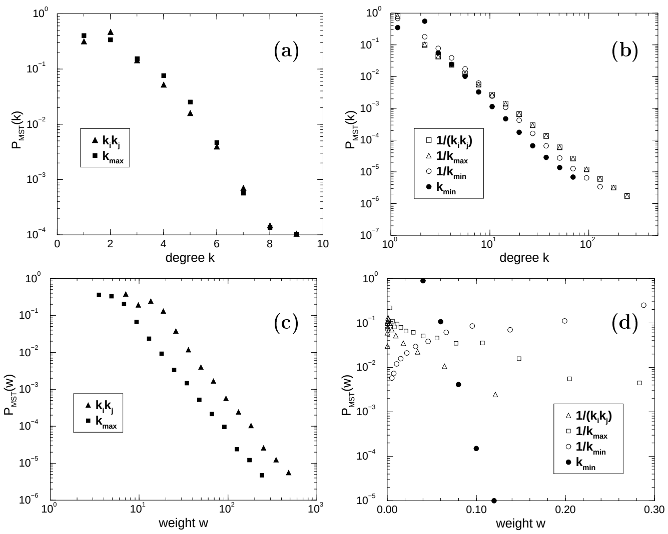

FIG. 3: Degree and weight distribution of $N = 10^4$ node MSTs. **(a)** The degree distribution for weights proportional to either k_ik_j (i) or k_{max} (ii) are dominated by an exponential cut-off, while **(b)** it is heavy-tailed for weights proportional to k_{min} (iii) and inversely proportional to either k_ik_j (iv), k_{max} (v) and k_{min} (vi). **(c)** The distribution of link weights on the MSTs is a power law for (i), (ii) and (vi) ($w \times 10^3$), while **(d)** it is dominated by an exponential cut-off for (iii) ($w \times 0.02$), (iv) and (v). For each weight choice we averaged over 10^4 different MSTs.

of the (i) and (ii) MSTs through the following argument: Since the MST tends to avoid links with large weights, it effectively shuns the hubs for the cases $w_{ij} = k_ik_j$ and $w_{ij} = k_{max}$, utilizing instead, whenever possible, links connecting low degree nodes (Fig. 3a). Consequently, all the hubs are marginalized and the MST degree distribution must have a narrow range. This argument is supported by Fig. 4a and b, where we show examples of MSTs for weight choices (i) and (ii) respectively. The sizes of the nodes in the figure reflect their degree in the original network. It is evident that the majority of the hubs are located on the branches ($k = 1$ degree nodes) of the MST (Fig. 4). This reliance on the small nodes and tendency to avoid the hubs forces the MSTs generated by method (i) and (ii) to be very similar to each other. Indeed, we find that for a given network but weights created by methods (i) and (ii), 87% of the links in the two MSTs are in common. This explains the similar visual appearance of the two MSTs (Fig. 4a and b).

The second class of MSTs is well represented by weight choices (iii)-(vi), resulting in power-law MST degree distributions. The similarity between weight schemes (iv)-(vi) is emphasized by the fact that their MST degree distributions follow a power-law with the same exponent $\gamma = 2.4$ [24] (Fig. 3b). Indeed, the links with the lowest weights are now connected to the hubs of the original network, and the MST grows utilizing these hubs extensively. Hence, the hubs of the full network experience only a slight reduction in their degree and are found at the center of the resulting MSTs (Fig. 4c), while the intermediate-degree nodes sustain large losses of neighbors and are found at the surface of the network with one or two neighbors (Fig. 4c).

The distribution of link weights on the MST also displays two distinct behaviors, being either power-law (Fig. 3c) or exponential (Fig. 3d). It is interesting to note that MSTs with exponential degree distribution (Fig. 3a) display power-law weight distributions with exponents $\sigma = 3.1$ (case (i)) and $\sigma = 3.0$ (case (ii)) (Fig. 3c). On the other hand, for weight choices (iii)-(v) the degree distribution of the MST is power law and the MST weight distribution is exponential or stretched exponential (Fig. 3d). For (vi) $w_{ij} = 1/k_{min}$ both the degree and the weight distribution of the MST are scale free. Finally, if the link weights are distributed uniformly and randomly the resulting MSTs have a power law degree distribution and an exponentially tempered weight distribution [31].

In order to investigate the effect of the degree correlations on the MSTs for weight choices (i)-(vi), we randomized the weights of the original network by randomly selecting pairs of links and exchanging their weights until all correlations between weights and the local network

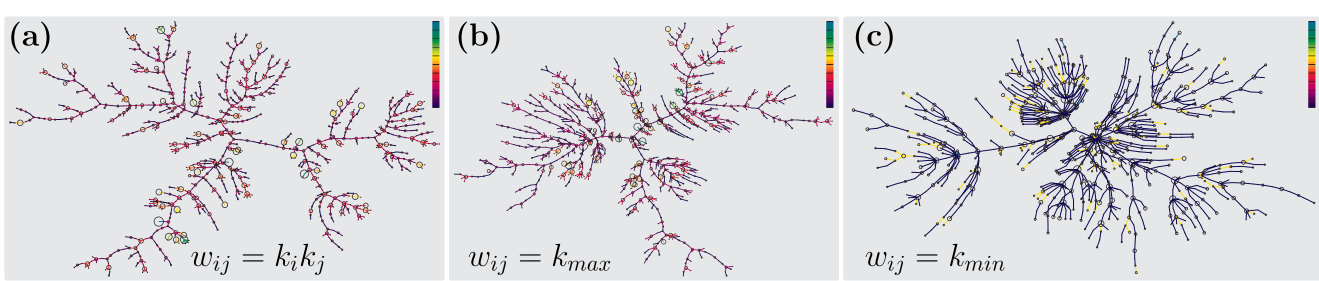

FIG. 4: (Color online) Minimum spanning trees of a $N = 10^3$ node scale-free network for weight choices **(a)** k_ik_j, **(b)** k_{max} and **(c)** k_{min}. The size of a node represents its degree in the full network, and the color of a link represents its weight from low (black) to high (green). Note that the MST degree distribution is exponential for **(a)** and **(b)** and a power law for **(c)**.

topology were lost. Invariably, the resulting MSTs were scale-free with a degree exponent similar to that of the original network, $\gamma \approx 3$. The local structure of the MSTs are very different, however, with only 52% of the links staying the same in a pairwise comparison between MSTs with weight choices (i) and (ii), suggesting that the functional form of the weight distribution is inconsequential for the degree distribution of the MSTs. To understand this we recall that only the *ranking* of the link weights and not their absolute value matters [23]. Therefore, by removing the correlations between the local network structure and weights we effectively map the problem onto that of weights being uniformly random. Indeed, the degree distribution of the MST in this case is also power-law with $\gamma \approx 3$ [31]. However, the MST weight distributions continue to depend on the weight distribution of the original network.

Discussion.—As networks play an increasing role in the exploration of complex systems, there is an imminent need to understand the interplay between network dynamics and topology. While focusing on the MSTs of scale-free networks, our results emphasize the significance of correlations between link-weights and local network structure. We find that if correlations are present, two classes of MSTs exist, following either a power-law or an exponential degree distribution. The removal of correlations renders the MSTs scale-free, independent of the choice of the weight distribution. This result raises interesting questions regarding our ability to quantify the influence of weights.

Our findings could serve as a natural starting point towards the systematic exploration of weighted networks. For example, while we have assumed that the weights are static, incorporating their time-dependence may reveal novel dynamical rules. Second, we model the weights as solely dependent on the topology, potentially overlooking correlations among the weights themselves. Uncovering the role of such correlations remains a challenge for future research.

We thank J. Kertész, P. L. Krapivsky and S. Havlin for discussions. We also thank M. A. de Menezes for sharing the US airport data. This work has been supported by grants from DoE (E.A.) and the R.E.U. program at Notre Dame (P.J.M).

* Electronic address: alb@nd.edu

[1] S. H. Strogatz, Nature **410**, 268 (2001); R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford Press, 2003); R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet : A Statistical Physics Approach* (Cambridge University Press, 2004); A.-L. Barabási and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101 (2004).
[2] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, Nature **407**, 651 (2000); D. A. Fell and A. Wagner, Nat. Biotechnol. **18**, 1121 (2000).
[3] H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai, Nature **411**, 41 (2001).
[4] R. Albert, H. Jeong, and A.-L. Barabási, Nature **401**, 130 (1999).
[5] H. Ebel, L.-I. Mielsch and S. Bornholdt, Phys. Rev. E **66**, 035103(R) (2002).
[6] R. Albert, H. Jeong and A.-L. Barabási, Nature **406**, 378 (2000); R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Phys. Rev. Lett. **85**, 4626 (2000); D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000); R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[7] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[8] S. H. Yook, H. Jeong, A.-L. Barabási and Y. Tu, Phys. Rev. Lett. **86**, 5835 (2001).
[9] K.-I. Goh, B. Kahng and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[10] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, H. E. Stanley, Phys. Rev. Lett. **91**, 168701 (2003).
[11] A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004).
[12] Z. Toroczkai and K. E. Bassler, Nature **428**, 716 (2004).
[13] M. Granovetter, Am. J. Soc. **78**, 1360 (1973).
[14] A. M. Kilpatrick and A. R. Ives, Nature **422**, 65 (2003).
[15] E. L. Berlow, Nature **398**, 330 (1999).
[16] R. M. Callaway, R. W. Brooker, P. Choler *et. al*, Nature **417**, 844 (2002).
[17] A.-L. Barabási, Phys. Rev. Lett **76**, 3750 (1996).
[18] G. B. West, J. H. Brown and B. J. Enquist, Science **276**, 122 (1997).
[19] J. R. Banavar, A. Maritan and A. Rinaldo, Nature **399**, 130 (1999).
[20] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan and A. Rinaldo, Phys. Rev. Lett. **84**, 4745 (2000).
[21] E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai and A.-L. Barabási, Nature **427**, 839 (2004).
[22] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[23] R. Dobrin and P. M. Duxbury, Phys. Rev. Lett. **86**, 5076 (2001).
[24] D.-H. Kim, J. D. Noh and H. Jeong, preprint cond-mat/0403719 (2004).
[25] R. V. Hogg and A. T. Craig, *Introduction to mathematical statistics*, 5th ed. New York, Macmillan (1995).
[26] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[27] Taking into account the degree-degree correlations of our model [26], we find the same $w \gg 1$ limiting behavior as that of Eqs. (1)-(3).
[28] A.-L. Barabási, M. A. de Menezes, S. Balensiefer and J. Brockman, Eur. Phys. J. B (in press), DOI: 10.1140/epjb/e2004-00022-4.
[29] R. C. Prim, Bell Syst. Tech. J. **36**, 1389 (1957).
[30] G. Szabó, M. Alava, and J. Kertész, Physica A 330, 31-36 (2003)
[31] The special case of the weights being uniformly random was recently published by Szabó et al. [30]. The MST of this network is scale-free with $\gamma \approx 3$, and our numerically determined weight distribution agrees with Ref.

[30]. The acceptance function $P_{MST}(w)/P(w)$ for the case of $w_{ij} = k_{max}$ and $w_{ij} = 1/k_{max}$ was obtained independently by Szabó, Alava, and Kertész. J. Kertész, *private communication.*